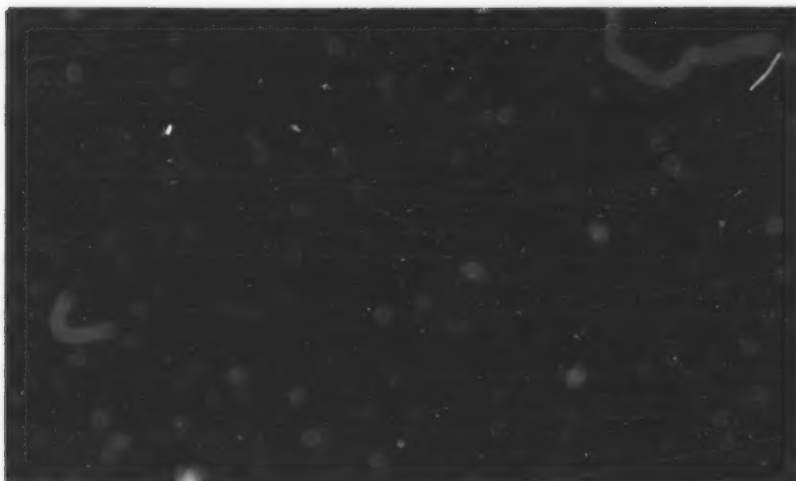




**CENTRE
DE RECHERCHE
SUR LES TRANSPORTS
(CRT)
CENTRE
FOR RESEARCH
ON TRANSPORTATION**

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



C.P. 6128, SUCCURSALE A
MONTRÉAL (QUÉBEC)
CANADA
H3C 3J7

TÉLÉPHONE : (514) 343-7575
TÉLEX : 05-24146 BIBPOLYTEC
TÉLÉCOPIEUR (FAX) : (514) 343-7121

Accepted in IEEE Transactions on Reliability.

CRT-548

COMMUNICATION AND TRANSPORTATION NETWORKS
RELIABILITY USING ROUTING MODELS

par

Brunilde Sanso¹
François Soumis²

¹ Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Québec CANADA H3C 3J7.

² École Polytechnique de Montréal, C.P. 6079, succursale A, Montréal, Québec CANADA H3C 3A4.

The authors are grateful to the Bell-Northern Research for their support.

Centre de recherche sur les transports - Publication # 548

August 1987 - revised April 1989 and June 1990

Dépôt légal - 2e trimestre 1990

ABSTRACT

A general framework for calculating a reliability measure for several types of flow networks is presented. This framework allows to perform reliability analysis for complex systems such as communication, electric power and transportation networks. The analysis is based on the notion of routing and re-routing after a failure. Different modelling approaches are discussed for every type of systems surveyed.

Key words : Reliability, Routing, Flow Networks, Optimization, Communication Networks, Transportation Networks.

- - - - -

RÉSUMÉ

Nous montrons une méthodologie générale pour calculer une mesure de fiabilité pour différents types de réseaux de flot. Telle méthodologie permet de faire des analyses de fiabilité pour des systèmes complexes tels que les réseaux de télécommunications, d'électricité et transport. L'analyse est basée sur la notion d'acheminement et re-acheminement du flot après panne. Plusieurs approches de modélisation sont discutées pour chaque système considéré.

Mots-clés : fiabilité; réseaux de flot; système de communication; réseaux de puissance; réseaux de transport.

1. INTRODUCTION

This paper aims at introducing a common procedure to evaluate reliability in diverse kind of flow networks. We will focus on new modelling approaches to find commonality among the different application areas.

Traditionally, two approaches have been taken to evaluate network reliability:

- 1 - One considers the network as a pure graph and, knowing node and arc reliabilities, evaluates some measure of connectivity as the reliability of the network. Many authors have studied and are still studying this kind of approach (see, for instance [1-5] and, more recently, the works compiled in [6]).
- 2 - One takes into account that the purpose of a flow network is to carry one or several commodities from origin to destination nodes in order to satisfy a certain demand. Several measures have been studied under this approach; in general, they are related to the ability of the network to transmit a level of flow [7-12].

Many of the measures of connectivity proposed in the first approach have been proven, or are suspected, to give rise to NP-hard problems [4]. Moreover, connectivity in most real networks is usually high even after important failures. Therefore, taking the probability of connectivity as a measure of reliability can lead to an overestimation of network performance. Furthermore, the first approach does not take into account the flow nature of the network, and related parameters such as arc capacity and demand are completely left out of consideration.

The clear advantage of the second approach is, as we have already mentioned, the fact that the flow nature of the network is specifically considered. However, both approaches ignore the fact that failures in real systems bring up the issues of routing.

By using the bounding method of Li and Silvester [13], we give more complex models that incorporate routing and show that they can be used in reliability evaluation. In other words, we propose a reliability measure that will take into account not only the network probability of failures or the capacity of the network to cope with its demand, but also the capacity of the system to adjust its flow after a failure.

The specific areas that will be surveyed in this work are: communication, electric power, and transportation (vehicular and air transportation). For some of these areas, reliability techniques have been widely applied to the study of measures of network performance (see [14, 15]). For others, such as transportation systems, traditional reliability approaches were difficult to model. We believe that our method can be considered useful and innovative for all areas treated.

In this paper we first outline the general methodology that will be later applied to evaluate the performance of specific networks. For every system, we then discuss the following: a mathematical description of the system, the measures of performance to be considered, a mathematical model of failures and, finally, routing models and (if available) routing algorithms.

2. GENERAL METHODOLOGY

The procedure for finding a good measure of performance in any of the network flow problems to be presented is summarized below. It is important to point out that the reliability parameter for most of the areas surveyed is an expected measure of performance and, therefore, bounds on the measure are to be found. However, for some systems, other kinds of statistics could be more interesting. The general methodology is then based on the evaluation of an expected measure; specific changes will be discussed within every system.

Methodology:

- 1 - Choose a performance measure Z such that the expected value of Z is the desired measure of reliability.
- 2 - Considering the kinds of failure which can occur in a network, partially enumerate the m most probable states of the network.
- 3 - For each one of the m states enumerated, find the routing and the value of Z for that specific scenario.
- 4 - Use the Z values computed for every state to obtain bounds on the expected value of Z .

2.1 Choice of Measure

First of all, we need to find a measure of performance that will consider the flow nature of the network. Two different groups of measures are immediately outlined, those which evaluate reliability from the point of view of the user and those which evaluate reliability from the point of view of the manager.

- 1 - **quality of service measures** (reliability from the user's point of view). A suitable quality measure could be related to the number and importance of customers served (this could be applied in

telecommunications and electric systems). In transportation, a quality of service measure is always related to time, for instance, mean time for a passenger to reach his destination or mean time of delay in the event of failure.

- 2 - **economic measures** (reliability from the manager's point of view). Failures can always be quantified in terms of money. Reliability can then be measured by loss of revenue in the case of failure.

2.2 Partial Enumeration

Let us define a state of the system as the state of every system component. If these components can only be failed or operating, there will be 2^n possible states (n being the number of components of the system). On the other hand, if all n devices of the system have s operating states, there will be s^n possible states. If each device i has a particular number $s(i)$ of states associated to it, the number of possible states is $\prod_{i=1}^n s(i)$. For the sake of simplicity, we refer to two-state devices; then in case a component can have several suitable and unsuitable levels of operation we have two choices:

- 1 - Consider a "failed state" as the union of all unsuitable levels and, consequently, the "operating state" will be the union of the suitable levels.
- 2 - Use an enumerating algorithm for multimode components such as the one described in [16].

Li and Silvester [13] have developed an algorithm to enumerate the m most probable states for two-mode component systems. However for systems whose arc availability is very high, no algorithm is really needed since the no-failure and single failure state will cover most of the space-state probability.

It is interesting to point out that for some systems, such as

vehicular transportation, the planner is not necessarily interested in the "m most probable states" but rather on the "m most important states", that could be the states that create the most damage in the system. In this case, neither the algorithm nor the bounds proposed in [13] are appropriate. We will discuss in more detail this situation when describing the vehicular transportation system.

2.3 Routing: Global and User's Optimization Notions

A general network can be formalized by a graph $G(N,A)$ where N represents the set of nodes and A the set of arcs. In a network, flows usually model some physical flow quantity such as flow of persons, electric power, fluid, or money. Nodes constitute points of origin, destination or transshipment of commodities. Arcs are direct paths between nodes.

Most physical network flow problems have to be modelled as multicommodity networks, that is, networks where the flow to be transported is of a varied nature, usually associated with the pair of origin and destination nodes of the flow (O/D pair).

We now present the notation necessary for constructing an optimal routing model for a multicommodity network. In the event that a single commodity network is considered, the formulation is still valid; it is only necessary to change the cardinality of the commodity set.

Notation:

x_{ij}^c : flow of commodity c flowing from node i to node j .

C_{ij} : capacity of the (i,j) arc.

d^c : demand for commodity c .

$\Gamma(i)$: set of arcs whose initial extremity is node i .

$\bar{\Gamma}(i)$: set of arcs whose terminal extremity is node i .

N : node set.

- A : arc set.
- N_C : commodity set.
- s^C : origin node for commodity c.
- t^C : destination node for commodity c.

When the routing procedure is a global optimization procedure taken from the manager's point of view it can be modelled by the following multicommodity problem:

$$\min L(x_{ij}^C) \quad (1)$$

s.t.

conservation of flow:

$$\sum_{(i,j) \in \Gamma(i)} x_{ij}^C - \sum_{(j,i) \in \Gamma(i)} x_{ji}^C = \begin{cases} d^C & \text{if } i=s^C \\ -d^C & \text{if } i=t^C \\ 0 & \text{otherwise} \end{cases}, \text{ for all } c \in N_C \quad (2)$$

capacity:

$$\sum_{c \in N_C} x_{ij}^C < C_{ij}, \quad \text{for all } (i,j) \in A \quad (3)$$

nonnegativity:

$$x_{ij}^C > 0, \quad \begin{matrix} \text{for all } (i,j) \in A, \\ \text{for all } c \in N_C. \end{matrix} \quad (4)$$

In (1), $L(x_{ij}^C)$ represents a linear function to be minimized which can simply be a dummy function, a valid function evaluating costs or minimum paths in the network, or even the value of the measure of reliability for the state considered.

The multicommodity problem with capacity constraints is very difficult to solve. Traditionally, the capacity constraints have been dropped and replaced by an artificial penalty function to be added to the objective function. In several of the routing applications we will review, the

penalty function is not artificial, but a carefully studied non-linear function that either corresponds to congestion delay or reflects the stochastic character of the demand while avoiding the capacity of the arcs to be surpassed. The problem can then be reformulated as:

$$\begin{array}{ll} \min & L(X_{ij}^C) + P(X_{ij}^C) \\ \text{s.t.} & \\ & \boxed{\text{conservation of flow}} \\ & \\ & \boxed{\text{nonnegativity}} \end{array} \quad (5)$$

where $P(X_{ij}^C)$ is the penalty function just described. Existing algorithms can solve, in reasonable time, large scale network problems formulated as (5).

The routing formulations shown up to now correspond to a global optimization model of the network. This kind of model considers that the network is managed by a sort of centralized decision maker who routes the flows in such a way as to optimize the global performance of the network. Global optimization is then a good approach to use in networks where no customers or users are involved in the outcome of the routing such as in electric power systems.

When users can decide the route to take, a decentralized behavior of the system has to be considered and the notion of "user optimization" has to be taken into account. A user optimization model can be a simple description of the equilibrium conditions of the system in order for the user to optimize a specific parameter. In general, a user optimization model can be formulated as a set of nonlinear equations as follows:

equilibrium conditions

flow conservation

(6)

nonnegativity

Some user optimization models are derived from the Kuhn-Tucker conditions applied to a network problem. For those models the possibility of solution by solving the corresponding network optimization problem exists (see Florian [17]). Otherwise, large scale user optimization formulations can be solved by methods such as Newton-Raphson and Gauss-Seidel or specialized methods for variational inequalities (see also [17]).

It is interesting to point out that global optimization routing gives a better performance measure than user optimization models. However, for some systems (such as vehicular transportation), only the users have control over the routing and, therefore, only user optimization models should be considered.

2.4 Performance Evaluation

When the number of states of a system is high, it is not practical to evaluate an exact expected measure of performance. In that case, bounds on the expected measure have to be computed.

Let S_k represent the k^{th} state of the system and Z the measure of performance to be considered. Then the upper and lower bounds on the expected measure of performance can be those proposed in [13].

$$E[Z]_{up} = \sum_{k=1}^m P(S_k) Z(S_k) + (1 - \sum_{k=1}^m P(S_k)) Z(S_{2^n})$$

$$E[Z]_{down} = \sum_{k=1}^m P(S_k) Z(S_k) + (1 - \sum_{k=1}^m P(S_k)) Z(S_1)$$

(7)

In (7), $P(S_k)$ represents the probability that the system is in the state S_k , $Z(S_1)$ is the measure of performance for the no-failure state and $Z(S_{2^n})$ is the outcome for the total failure state (all components have failed). $Z(S_1)$ and $Z(S_{2^n})$ must be such that $Z(S_1) = \min_k Z(S_k)$ and $Z(S_{2^n}) = \max_k Z(S_k)$.

The principal appeal of the bounds is based on the fact that they can be used with the re-routing option in mind. m routing problems have to be solved but, since they are very similar to one another, we will only need to re-optimize the routing of the $(k-1)^{th}$ state in order to find $Z(S_k)$.

We do not put great importance on the closeness of the bounds. In fact, since they are based on the outcome of the routing, their reference point could vary widely depending on the routing policy considered. In general, for "pessimistic" measures of reliability we will be interested in the upper bound. Sometimes, when the most damaging states are considered, our interest will not even be in the bounds but rather in the contribution of those states to the value of performance.

We now discuss the influence of routing in performance evaluation. In general, re-routing the demand after a failure will give better results in terms of performance than losing the demand when a component fails.

However, some physical networks do not have the capacity to re-route flow. Carey and Hendrickson [11] show excellent bounds on expected performance measures when no re-route is done.

3. MEASURES OF RELIABILITY ON SPECIFIC NETWORKS

3.1 Communications

The System

In a communication network, switching stations correspond to nodes and communication links to arcs. Flow in this type of networks can be measured in a unit of traffic called the Erlang (or CCS in the American system). Messages going from an origin to a destination node constitutes a commodity. This is a multicommodity system having as many types of flows as origin-destination pairs of nodes.

Networks of this kind can be further classified into two general types: circuit switching and packet switching. The basic difference between the two types is the strategy considered to allocate a request of call. If all the necessary resources that are required for the complete duration of the connection are reserved once a call is requested, we have circuit switching. For calls with low utilization factor, the allocated resources are used only a short period of time. In that case, circuit switching constitutes an inefficient way to route the calls. For these types of calls the suitable switching is packet switching. In these networks, the information is divided into "packets", and the resources to transmit a call are then only partially allocated from one switching station to the other for the time it takes a packet to travel between the stations. In general, telephone calls use circuit switching networks while packet switching is used to transmit data (computer networks).

Since the nature and the behavior of the flow depends on the type of switching, the modelling will also change accordingly. In fact, circuit

switched networks do not allow delay: calls are simply routed or lost. On the other hand the notion of overflow (lost calls) is not present as such in packet switching but transmission delays play a very important role in the routing.

In this paper, we will specifically model circuit switched networks. However, since most of the work in telecommunication networks reliability was done for packet switching, it is worth to mention the pioneering work by Frank and his colleagues in this field [18-24].

One of the very first studies on circuit-switching was done by Lee [25]. In that paper, methods for finding the blocking probability of a network were elaborated. In a sense the blocking probability of a network can be seen as a measure of reliability where failures model the event that a call is blocked. This idea has been studied by several authors in the circuit-switching field (see chapter 4 of Girard [26] and its numerous references). This type of approach focuses on the evaluation of system performance considering only the capacity of the equipment.

In our approach the capacity of the equipment, the damage caused by the failures as well as the ability of the system to adjust to failure are considered. A good measure of reliability that would take into account all the features we have discussed is the expected lost call traffic in the system. A methodology to evaluate this measure has already been developed in [27] and [28].

Failures

In communication networks, both nodes and arcs can fail. If n is the number of arcs and s the number of nodes, there will be 2^{n+s} states to be enumerated. A partial enumeration procedure to find the m most probable states should not discriminate between the kinds of devices which produce a failed state. Therefore, switching stations and communication links can be

treated as simple devices while finding the m most probable states. On the other hand, once placed in the routing subproblem, a distinction has to be made regarding the failure of arcs and nodes. In fact, when a node fails some or all the arcs adjacent to it fail as well. We have then considered that if a state admits the failure of some nodes the corresponding routing will take place in a subgraph where the failed arcs as well as the failed nodes and their adjacent arcs will be removed.

Another problem regarding failures arises in communication networks. Usually the communication company works with a logical network in mind and failures are referred to that kind of network. Unfortunately, since most of the time, different logical paths share the same physical equipment, failures in the logical network can be dependent. The algorithm of partial enumeration considers that device failures are independent. In order to overcome this difficulty, we can choose to work on the physical network where failures are independent, and translate the necessary results into the logical one.

Routing

In [27] and [23] several global optimization models were presented. The basic idea was to use the nonlinear formulation of a multicommodity problem given by (5). In this case the penalty function $P(x_{ij}^c)$ chosen to avoid the capacity of the arcs to be surpassed and, at the same time, to reflect the stochastic character of the demand was Erlang's expected loss call traffic function. The formulation for this kind of model was then:

$$\min \quad L(x_{ij}^c) + \sum_c \sum_{(i,j)} x_{ij}^c B\left(\sum_c \sum_{(i,j)} x_{ij}^c, c_{ij}\right)$$

s.t.

conservation of flow

(8)

nonnegativity

In (8) $L(x_{ij}^c)$ is a term to favor the shortest path and B corresponds to Erlang's loss formula, that is, the function that gives the probability of blocking in a link. To simplify the notation, let s be the capacity of the link (in number of circuits) and a the demand on the link (or offered traffic) in units of flow (Erlangs). Then:

$$B(s,a) = \frac{a^s/s!}{\sum_{i=1}^s a^i/i!} \quad (9)$$

The number of variables in (8) can be made smaller by considering one commodity as all the flows having the same origin.

We now show another type of routing model related to the notion of user optimization.

Although global optimization models give better performance results than user optimization models, actual present-day switching policies are closer to decentralized routing.

In the decentralized model we propose that every origin-destination pair will have a choice of three routes. The demand attempts to pass through the first route, the overflow from the first route attempts the second route, the overflow from the second route attempts the third route and the overflow from the third route will be lost traffic. Figure 1 shows this flow behavior.

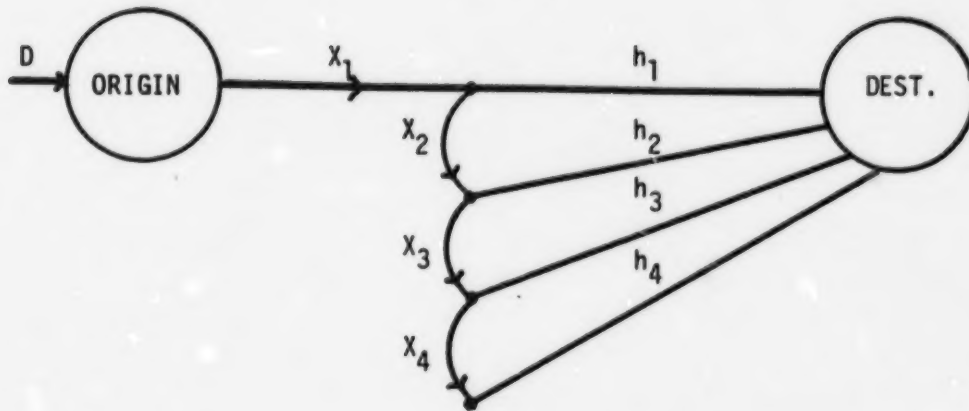


Figure 1

Let

x_q : average traffic trying to pass through route q .

h_q : average traffic flowing through route q .

$P_q(x_q)$: expected lost call traffic of route q .

D : demand for an O/D pair.

Then,

$$x_q = h_q + P_q(x_q), \text{ for all } q=1, \dots, 4 \quad (10)$$

In (10) we have introduced a dummy route (the fourth one) which receives the total traffic lost for the O/D pair.

Therefore

$$\begin{aligned} x_1 &= D \\ x_2 &= P_1(x_1) \\ x_3 &= P_2(x_2) \\ x_4 &= P_3(x_3) \end{aligned} \quad (11)$$

(11) corresponds to a set of equations that model the behavior for one O/D pair. When all origin-destination pairs are considered, (11) becomes a more complex equation system since several O/D routes can have arcs in common. Therefore, the excess flow in an O/D route depends not only on the flow from this very route, but from all flows of the different O/D routes that use the common arcs in the route. Then we can rewrite (11) as:

$$X_{i+1} = P_i(X_i, x), \quad i=1,2,\dots,3 \quad (12)$$

where x is a function englobing the flow variables in the common arcs. The equation system obtained by applying (12) for all O/D pairs can be solved by numerical methods such as Gauss-Seidel or Newton-Raphson.

Regarding an expression for $P_i(X_i, x)$, we refer the interested reader to [28].

Reliability Evaluation

In real telecommunication networks, components are extremely reliable. If only link failures are considered, since link reliability is greater than 99 %, single failure and no-failure states will cover most of the state space probability.

Then, for each of the states, a routing problem will be solved. If the chosen model is a global optimization one, the function to be evaluated for every state is:

$$Z(S_k) = \sum_c \sum_{(i,j)} X_{ij}^c B(\sum_c \sum_{(i,j)} X_{ij}^c, C_{ij}) \quad (13)$$

instead, for the decentralized model:

$$Z(S_k) = \sum_c X_4^c \quad (14)$$

where X_4^c is the total overflow for commodity c .

Finally it is interesting to point out that even though models have been developed for circuit switched networks, similar models could be presented for packet-switching. In that case, the measure of reliability should be related to the expected delay in the network and the penalty function in (5) would be the delay function for a link. Due to the flexibility of packet switching, global optimization models could give results much closer to reality for those networks than for circuit-switched ones.

3.2 Electric Power Systems

The System

A complete power system is composed of generation, transmission, and distribution facilities. The aim of the system is to meet the power requirements of the changing load. Even though power system consumers are usually familiar with the notion of "energy", the power company is concerned instead with the needs of "power" in the system, that is the instantaneous rate of energy variation. With this idea in mind, the power system can be modeled as a flow network where the commodity to be transmitted is the electric power. Generally speaking, the term power systems refers to AC power systems (that is systems where the power, the load and the line parameters are modeled as complex variables).

Traditionally, power system reliability was studied in terms of indices such as the LOLP (loss of load probability) or LOLE (loss of load expectation) that showed the ability of the system to cope with specific levels of demand [15]. In order to determine the reliability indices, the entire system was not considered; the system was rather decomposed into suitable subsystems and the ability to satisfy customer load requirements was studied within each subsystem.

In our approach we will consider the system as a whole. The network is modelled as a graph $G(N,A)$ where N , the set of nodes, represents generating stations, switching stations, or load bars, and A , the set of arcs, represents the transmission lines.

Although we are really concerned with AC power systems, it is interesting to mention the work by Wu and his colleagues [29-32] in DC power system reliability evaluation since it presents some common points with ours. In general, they have considered the system as a whole and have modelled it as a probabilistic graph. A DC load flow study is implicitly or explicitly considered. We will explain the characteristics of a load flow study later in this section, however we mention that its nature is related to the routing of commodities in the system and, therefore, with the flow nature of the network. In [29] and [30] the reliability of a DC power system is defined as the probability of the configurations for which the DC load flow is satisfied with a specific level of load demands. The solution approach does not require the solution of the load flow although it implicitly considers the satisfaction of the load flow. In [31] the authors introduce a new reliability index, the inadequate transfer capability from two areas of the network. The solution method requires a space state decomposition and a Monte Carlo simulation. Liu and Wu [32] show how to model an electric power supply-demand problem as a compact flow network and derive some conditions under which this model can be used in reliability evaluation.

We propose two different kinds of measures to evaluate the network performance. In fact, for this kind of system measures of reliability from the user's as well as the operating company's point of view can be chosen. In either case the measure will be expected values or weighted expected values calculated by carrying out the partial enumeration procedure and the routing for every state.

A measure of service quality could be the expected number of times a specific customer or region is left without service or the weighted expected number of customers shut off in the study period.

From the company's point of view, economic measures such as expected generation costs or expected loss of revenue due to failures may be taken into account.

Failures

For the system under study, failures of both nodes and arcs should be considered. Since a node represents a generating station, a failure can model a reduction on the generation capacity. Arc failures represent the physical interruption of the flow on the failed transmission line.

In AC systems, the generation and the transmission is done in a triphasic way. However, when no failures are present, load flow studies are carried out by considering monophasic buses and lines that carry triphasic parameters. Depending on the type of failure (monophasic, biphasic, triphasic, etc.) a different scenario of overcurrents will take place. Our model does not study those scenarios. We consider instead that after any kind of failure the circuit breakers that protect part of the system will be opened and will isolate one or several components in the system. Therefore, if the failed element is a generator the failure will be modelled as a reduction in the generation capacity and, eventually, a loss of one or several arcs. On the other hand a transmission line failure is represented by the loss of one or more arcs in the system.

Finally, it is important to point out that transitory effects due to the failures, before and after the opening of breakers, are not taken into account.

Routing

Traditionally, power engineers have routed power flows using a method called a load flow study. This method consists in a set of flow conservation equations (the flow being a complex variable in AC systems) and inequalities derived from stability requirements of the network. Since the number of variables is greater than the number of equations, some variables are defined as dependent and others as independent (see Elgerd [33]). Once the independent variables have been established, a Gauss-Seidel or Newton-Raphson algorithm determines the value of the dependent variables. The operator will then check if the inequality constraints are satisfied, if not, it is up to his knowledge and ability to adjust parameters and rerun the program. There are, therefore, many possible solutions to the same load flow study.

The goal of much research since the early 1960's has been to find the optimal load flow study, the term "optimal" generally referring to economic criteria. The problem is then formulated as:

$$\begin{aligned} &\min f(Z) \\ &\text{s.t.} \\ &g(Z) = 0 \\ &h(Z) < 0 \end{aligned} \tag{15}$$

where Z is the vector of the variables of the system, $g(Z)$ represents the flow conservation equations and $h(Z)$ the stability and "physical" constraints. This problem is not easy to solve due to the complex nature of the variables and the non-linearity of both the objective function and the constraints. In general, some approximations had to be made by dividing the problem into two subproblems: first, a subproblem treating the real power and the variables dependent to it; and second, a subproblem treating the reactive power and its dependent variables.

In general the optimal flow studies need a feasible starting solution and lack global convergence (see [34]). In [34] a new method is proposed which does not need a feasible starting solution, does present global convergence and is easy to deploy.

Reliability Evaluation

For power systems we can consider the most probable states and evaluate bounds on expected reliability measures or just take into account some strategically important states and evaluate their contribution to the expected measures of reliability.

As for the measures to be considered, if the idea is to evaluate quality of service, a load flow study can determine if a customer i has been left without service. With this knowledge either the probability that a specific customer will be out of service or the expected number of times in the study period that the customer is shut off could be evaluated. If reliability is measured in economic terms, then the value $Z(S_k)$ to be considered could be the generation costs $f(Z)$ as in (15) or another more complete function which could evaluate the benefit of the company taking into account generation costs and expected revenue in the event of failures.

3.3 Vehicular traffic

The System and the Routing

In a formal context, a vehicular traffic transportation network is a flow network where nodes represent origins, destinations or intersections; links represent the transportation infrastructure and the commodities to be transported are the flow of vehicles travelling from origins to destinations.

The routing in the system, known in the transportation field as traffic assignment, can be modelled as a network equilibrium problem (see

Florian [17]). This kind of model describes the behavior of the users from their origins to their destinations so that link flows and travel times can be estimated. Unlike other kinds of systems in which global optimization can reflect reality, users of the vehicular transportation system behave in a selfish way; therefore user's optimization models must be those considered. The assumptions for users' behavior were stated by Wardrop [35]. They implied that the routes actually used are the shortest, in terms of travelling time, under the prevailing traffic conditions.

Let:

A : set of links.

v : vector of link flows.

v_a : flow on link a .

$s_a(v)$: cost of travelling on link a .

I : set of all origin-destination pairs.

K_i : set of paths for origin-destination $i \in I$.

g_i : demand for O/D i .

h_k : flow on path k .

Then the equations describing user's optimization behavior are:

$$s_k^* - u_i^* \begin{cases} = 0 & \text{if } h_k^* > 0 \\ > 0 & \text{if } h_k^* = 0 \end{cases} \quad k \in K_i, i \in I, \quad (16)$$

over (17) to (21),

$$\sum_{k \in K_i} h_k = g_i, \quad i \in I \quad (17)$$

$$h_k > 0, \quad k \in K_i, \quad i \in I$$

where

$$\delta_{ak} = \begin{cases} 1 & \text{if link } a \text{ belongs to path } k \\ 0 & \text{otherwise,} \end{cases}$$

$$s_k = \sum_{a \in A} \delta_{ak} s_a(v), \quad k \in K_i, i \in I \quad (19)$$

$$u_i = \min_{k \in K_i} s_k, \quad i \in I \quad (20)$$

and g_i being a function of u , the vector of the least cost travel times:

$$g_i = G_i(u), \quad i \in I \quad (21)$$

Expressions (16) to (21) state the equilibrium conditions of the system. For some cases (16) to (21) correspond to the Kuhn-Tucker conditions of an optimization problem and non-linear programming algorithms are used to solve it (see Le Blanc et al. [36]; Florian and Nguyen [37]). In the more-general case (16) to (21) are solved as a variational inequalities problem (see Aashtiani and Magnanti [38] and Nguyen and Dupuis [39]).

The Failures and the Model

As we have seen, the equilibrium problem is highly "static"; a network equilibrium solution can indicate, for instance, the best route from an origin to a destination under "normal" or specific conditions. However, if in this route there is a high probability of accidents, it is possible that the route will not be so convenient. The idea is then to incorporate the event of accidents into the equilibrium models.

An accident can be modelled as a failure and every type of failure can be represented as a specific diminution of arc capacity. To simplify the model only specific levels of link capacities should be considered.

Now if the general methodology were applied without any changes, that is for every failure state an equilibrium problem is solved with reduced capacity in the failed link, that would mean that before departing the drivers would already know that an accident will take place. This does not correspond to reality. However, this kind of model could be useful to study traffic behavior when link capacity reductions have been announced (for instance in case of programmed road improvements or repairs). The problem we are faced with is to model a highly dynamic behavior, as is the one by the drivers after an accident, with a static tool, as is the classic equilibrium algorithm.

Without changing the equilibrium algorithm we will propose a three stage model that will attempt to simulate the dynamic of the accident. Figure 2 shows the partition of the temporal axis when an accident takes place.

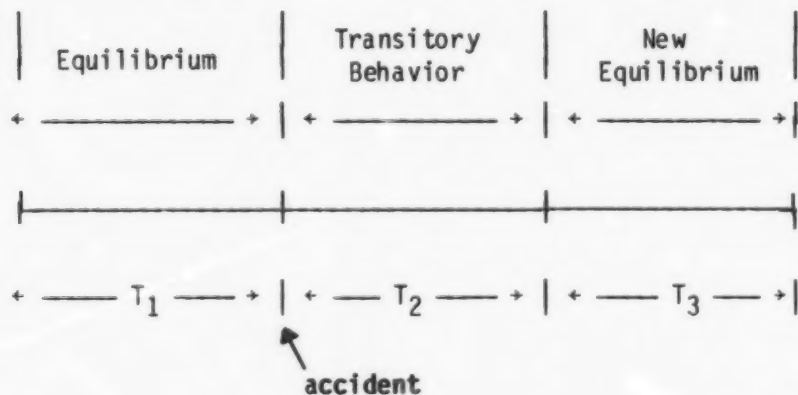


Figure 2: Temporal axis corresponding to the dynamics of the accident

During period T_1 the system is in the no-failure state and has reached the equilibrium. Once the accident takes place a bottleneck will be formed in the failed link due to the reduction of its arc capacity. A new state of duration T_2 is entered. During this period of time, several phenomena will be observed:

- 1- the routing (traffic assignment on every link) will not change;
- 2- on the routes using the link where the accident occurred, there will be accumulated flow on the links preceding the link on which the accident took place and volume reduction on the links following that link;
- 3- on those links where the traffic volume has increased there will be an increase in travel time and where the traffic volume has decreased travel time will be reduced.

After a while, in a state of period T_3 , some users will have been informed of the accident. Those drivers will avoid the link where the accident occurred by using an alternate route. An assumption has to be made regarding the percentage of the demand that, at the origin, is informed of the accident. The situation for this period will be modelled by removing the assumed percentage of the informed demand from the links and re-assigned it using an equilibrium algorithm.

Reliability Evaluation

It has already been discussed how for a vehicular traffic system it is impractical to enumerate the most probable states. Instead, the states that consider the most important accidents should be studied. Then, reliability evaluation could be done by computing the contribution of these states to the measure of performance. For instance, if the system under study corresponds to a city whose physical location is an island, an accident on a bridge should be a major concern. Accidents on main highways should also be considered.

In order to evaluate the probability of entering a state, statistics should be available, if not those probabilities could be assumed and used as parameters. It is important to point out that every failure state considered under the three-period-model is itself composed of two "sub-states". The probability of being on any sub-state will depend on its duration.

With regards to the measure of reliability to be studied, statistics such as the frequency of the mean travel time for every origin-destination pair as well as the variance of the mean travel time can be relevant measures.

3.4 Air Transportation

The System

The air system can be represented by a space-time network where each node consists in a place and a time of arrival or departure from that place. In the more general form of this network, arcs represent segments of flights between cities.

The flows in the network are the passengers and the type of flow will be determined by the origin-destination points associated with each passenger. Before getting into any other details of the system, we introduce some suitable definitions (extracted from [40]).

- Flight : service description. A flight is defined by a plane, an itinerary (Montreal-Rio-Buenos Aires), departure and arrival time.
- Flight Segment : nonstop segment of a flight (Montreal-Rio).
- Route : path from an origin to a destination used by the passengers.
- Market : group of passengers wanting to travel from a specific origin to a specific destination within a time interval.

With this basic terminology in hand, we can now describe the network more in detail. There are three kinds of arcs to be considered:

- 1) Arcs defined as flight segments.

- 2) Arcs representing airport waiting time for a scheduled stop.
- 3) Arcs representing airport waiting time between the arrival of a flight and a connecting flight departure.

For this type of system, measures of reliability from both the user's and company's point of view can be considered.

A good measure of reliability from the company's point of view could be the expected number of lost customers due to failures in the study period. This measure could be weighted according to the importance of the market and could be translated into expected loss of revenues in the event of failure.

If we assume that all passengers of a specific market have a choice (in a pre-specified priority) of routes, a measure of reliability from the passenger's point of view could be the expected priority number of the route finally used for the passenger. Another measure could be the expected increase in time of travel due to failures.

Failures

A failure in an air transportation network can be characterized by the absence of one or more arcs in the network. For instance, a delay due to bad weather or unscheduled airplane repairs can be considered as a failure. When a failure occurs before the beginning of the flight, not one but all arcs belonging to that flight must be removed from the network and eventually placed later in time. If the failure takes place in the middle of a flight, all arcs from the specific stop where the failure occurred to the end of the flight should be removed. In any case, after every kind of failure not only flight segments and scheduled stop arcs must be removed but it has also to be verified if connecting flight arcs are still feasible. Therefore, the failure scenario for this type of networks is more complicated than those considered for other systems.

Routing and Reliability Evaluation

Soumis [41] has developed four routing models in air transportation: two correspond to the idea of global optimization, and the other two model the system from the user's point of view.

In Figure 3 passengers' behavior according to one of Soumis' user's optimization models is depicted. In the model an attraction coefficient a_i is given to every possible route for the market. Potential passengers are assigned on the routes in proportion to the given attraction coefficients. A stochastic function which has the capacity of the links in the route as a parameter will evaluate the expected number of refused passenger in every route. Some of the passengers that did not find a place in the route of their choice will choose not to travel, the others will try again. An assumption has to be made regarding the proportion of the demand that chooses not to travel.

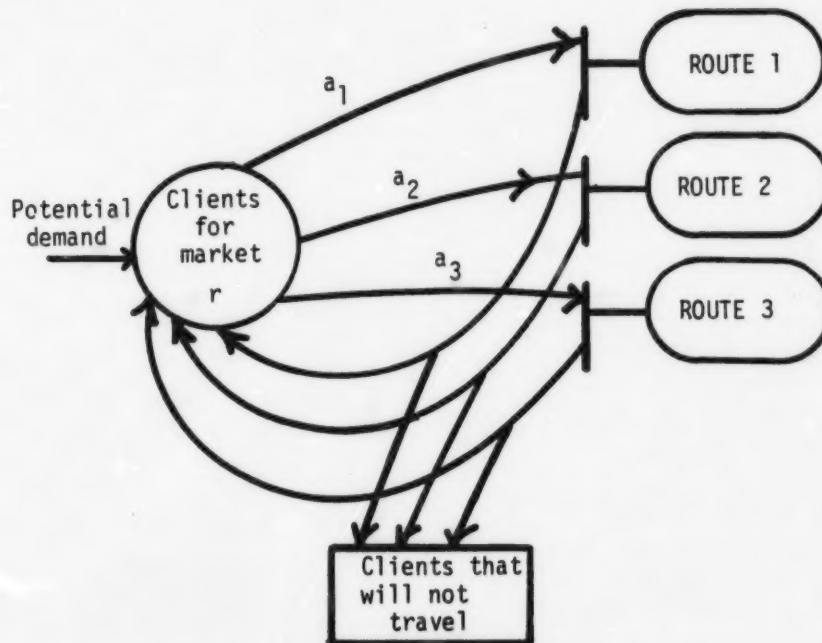


Figure 3: Passenger's behavior according to a user's optimization model

In normal circumstances the assignment of passengers is carried out according to airplane capacity. Some potential clients are not served, however, because they have not been assigned to a suitable route or because they have not found any available route. This "unserved demand" represents, in monetary terms, a loss of revenue for the airborne company. This loss is due not to any kind of failure but to the company's planning of routes and airplane capacity. This very situation is aggravated when a failure takes place; in fact, in this case the space-time network will be re-arranged and the choice of routes will be much smaller. Therefore, there will be a number of potential passengers which, at the departure node, will choose not to travel. With this idea in mind, a measure of reliability as a loss of revenue would be evaluated by considering all failure scenarios in a period of time and computing the loss of revenue for every scenario.

From the passenger's point of view, and under the behavior described by Figure 3, a measure such as the expected travel time or the variance of the travel time in the event of failure can be considered. Again, for every failure scenario there will be a different choice of routes. Then, within every state the maximum, minimum and mean time of travel could be measured for every market. Once these measures are known for every state, the desired statistics can be computed for every market as well as for the system as a whole.

4. CONCLUSIONS AND FURTHER WORK

This paper has established a general framework for calculating a reliability measure in different systems whose common characteristic is to carry flow. The proposed methodology underlines the importance of the routing in the reliability of flow networks. In fact, the performance of the system is not only based on the physical characteristics of the network but rather on the ability of the network to adjust to failures. This has been a specific focus of the paper.

Within every system we have proposed several different measures to evaluate network performance. For those systems for which reliability has been widely studied, the introduction of the routing procedure makes the modelling of the system closer to reality.

In transportation systems, the notion of reliability related to the ability of the system to re-route its flow is, to our knowledge, introduced for the first time. The measures proposed for vehicular as well as air transportation systems, could be useful not only to evaluate system performance but also to detect the user's need to change modes (car to bus, airplane to train, one air company to another) before it actually happens.

Regarding further work, we are currently working on two application fields: telecommunication networks design under reliability constraints and vehicular traffic. In the latest field the three-period model is being refined and the possibility of a dynamic traffic assignment model is being considered. Another area of application that is being explored is freight transportation. For electric power systems and air transportation we have provided the idea of a new methodology to evaluate reliability, but much work needs to be done by experts in every field to adapt these new ideas to their specific needs.

ACKNOWLEDGMENTS

The authors wish to thank David Scott and Professor Benjamin Smith for careful readings of the manuscript and the anonymous referees for their valuable suggestions.

REFERENCES

- [1] VAN SLYKE, R., FRANK, H. "Network Reliability Analysis: Part I", Networks 1, 279-290 (1972).
- [2] FRATTA, J., MONTANARI, U.G. "A Boolean Algebra Method for Computing the Terminal Reliability in a Communication Network", IEEE Transactions on Circuit Theory CT-20(3), 203-211, May 1973.
- [3] HANSLER, E., McAULIFFE, G.K., WILKOV, R.S., "Exact Calculation of computer Network Reliability", Networks 4, 95-112 (1974).
- [4] BALL, M.O. "Computing Network Reliability", Operations Research 27(4), 823-838, July-August 1979.
- [5] PROVAN, S., BALL, M.O. "Computing Network Reliability in Time Polynomial in the Number of Cuts", Operations Research 2(3), 516-526, May-June 1984.
- [6] IEEE Transactions on Reliability R-35(3), August 1986.
- [7] DOULLIEZ, P., JAMOULLE, E. "Transportation Networks with Random Arc Capacities", Revue Française d'Automatique, Informatique et Recherche Opérationnelle, November 1972.
- [8] LEE, S.H. "Reliability Evaluation of a Flow Network", IEEE Transactions on Reliability R-29(1), 24-26, April 1980.
- [9] AGGARWAL, K.K., CHOPRA, Y.C., BAJWA, J.S. "Capacity Consideration in Reliability Analysis of Communication systems", Transactions on Reliability R-31, 177-181, June 1982.
- [10] SHOGAN, A.W. "Modular Decomposition and Reliability Computation in Stochastic Transportation Networks Having Cutnodes", Networks 12, 255-275 (1982).

- [11] CAREY, M., HENDRICKSON, C. "Bounds on Expected Performance of Networks with Links Subject to Failure", Networks 14, 439-456 (1984).
- [12] AGGARWAL, K.K. "Integration of Reliability and Capacity in Performance Measure of a Telecommunication Network", IEEE Transactions on Reliability R-34(2), 184-186 (1985).
- [13] LI, V.O.K., SILVESTER, J.A. Performance Analysis of Networks with Unreliable Components", IEEE Transactions on Communications COM-32 (10), 1105-1110, October 1984.
- [14] FRANK, H., FRISCH, T. Communication, Transmission and Transportation Networks, Addison-Wesley (1971).
- [15] DHILLON, B.S., SINGH, C. Engineering Reliability, New Techniques and Applications, John Wiley & Sons (1981).
- [16] GAEBLER, R., CHEN, R. "Enumerating the States of a Communication Network Facility", working paper presented at the ORSA-TIMS Conference, St-Louis, Missouri, October 25-28, 1987.
- [17] FLORIAN, M. "Nonlinear Cost Network Models in Transportation Analysis", Mathematical Programming Study 26, 167-196, North-Holland (1986).
- [18] FRANK, H., FRISCH, T., CHOW, W. "Topological Considerations in the Design of ARPA Computer Network", AFIPS Conference Proceedings V36, 581-587, Atlantic City, New Jersey, May 5-7, 1970.
- [19] FRANK, H., FRISCH, T., CHOW, W. "Topological Optimization of Computer Networks", Proceedings of IEEE 61, 1385-1387, November 1972.
- [20] FRANK, H., "Providing Reliable Networks with Unreliable Components", IEEE 3rd Data Communications Symposium, 162-164, St-Petersburg, Floride, November 13-15, 1973.

- [21] FRANK, H., VAN SLYKE, R. "Reliability Considerations in the Growth of Computer Communication Networks", 1973 National Telecommunications Conference II, 22D/1-5, Atlanta, Georgie, November 26-28, 1973.
- [22] FRANK, H., GERLA, M., CHOW, W. "Issues in the Design of Large Distributed Communication Networks", National Telecommunications Conference, vol. II, Atlanta, Georgie, November 26-28, 1973.
- [23] FRANK, H., CHOW, W. "Network Properties of ARPA Computer", Networks 4(3), 213-239 (1974).
- [24] BALL, M., VAN SLYKE, R.M., GITMAN, I., FRANK, H. "Reliability of Packet Switching Broadcast Radio Networks", IEEE Transactions on Circuits and Systems 23(12), 806-813, December 1976.
- [25] LEE, C.Y. "Analysis of Switching Networks", Bell System Technical Journal 34(6), November 1955.
- [26] GIRARD, A. Routing and Dimensioning in Circuit Switched Networks, Addison-Wesley (forthcoming).
- [27] SANSO, B., GENDREAU, M., SOUMIS, F. "On the Evaluation of Telecommunications Network Reliability Using Routing Models", publication #623, Centre de recherche sur les transports, Université de Montréal, February 1988.
- [28] SANSO, B. "Fiabilité et routage dans un réseau de télécommunications" Ph.D. thesis, publication #605, Centre de recherche sur les transports, Université de Montréal, October 1988.
- [29] MOLESKI, K., WU, F.F. "Direct Method for Evaluation of Bulk Power System Reliability. Part 1: Theoretical Foundation", Electrical Power and Energy Systems 5(1), 3-7, January 1983.

- [30] MOLESKI, K., WU, F.F. "Direct Method for Evaluation of Bulk Power System Reliability. Part 2: Solution Algorithm", Electrical Power and Energy Systems 5(1), 8-16, January 1983.
- [31] CHANCY, D.P., GROSS, G., WU, F.F. "Probabilistic Flows for Reliability Evaluation of Multiarea Power System Interconnections", Electrical Power and Energy Systems 5(2), 101-114, April 1983.
- [32] LIU, C.C., WU, F.F. "A Generalized Network Flow Model with Application to Power Supply-Demand Problems", Networks 14, 117-139 (1984).
- [33] ELGERD, O.I. Electric Energy Systems Theory: An Introduction, McGraw Hill, 3rd edition (1975).
- [34] LOPES, L. "Le problème de répartition optimale des puissances pour les systèmes d'énergie en régime permanent", Ph.D. Thesis, département d'Informatique et de Recherche opérationnelle, Université de Montréal (1989).
- [35] WARDROP, J. "Some Theoretical Aspects of Road Traffic Research", Proceedings of the Institute of Civil Engineers. Part II, vol. I, 325-378 (1952).
- [36] LeBLANC, L.J., MORLOK, E.K., PIERSKALLA, W.P. "An Efficient Algorithms to Solve the Road Network Equilibrium Traffic Assignment Problem", Transportation Research 5, 309-318 (1975).
- [37] FLORIAN, M., NGUYEN, S. "An Application and Validation of Equilibrium Trip Assignment Methods", Transportation Science 10, 374-389 (1976).
- [38] AASHTIANI, H., MAGNANTI, T. "A Linearization and Decomposition Algorithm for Computing Urban Traffic Equilibrium", Proceedings of the 1982 IEEE Large Scale Systems Symposium, October 1982.

- [39] NGUYEN, S., DUPUIS, C. "An Efficient Method for Computing Traffic Equilibria in Network with Asymmetric Transportation Costs", Transportation Science 18, 170-191 (1984).
- [40] LAVOIE, S. "Modèles de comportement de passagers en transport aérien", publication #408, Centre de recherche sur les transports, Université de Montréal (1984).
- [41] SOUMIS, F. "Planification d'une flotte d'avions", publication #133, Centre de recherche sur les transports, Université de Montréal (1978).